

JACQUES BERTIN

La cartographie statistique automatique

Mathématiques et sciences humaines, tome 17 (1966), p. 71-76

[<http://www.numdam.org/item?id=MSH_1966__17__71_0>](http://www.numdam.org/item?id=MSH_1966__17__71_0)

© Centre d'analyse et de mathématiques sociales de l'EHESS, 1966, tous droits réservés.

L'accès aux archives de la revue « Mathématiques et sciences humaines » (<http://msh.revues.org/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

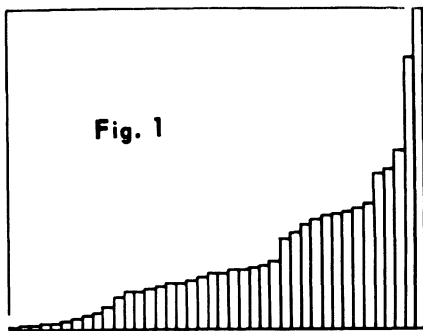


Fig. 1

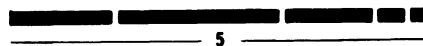


Fig. 2



Fig. 3

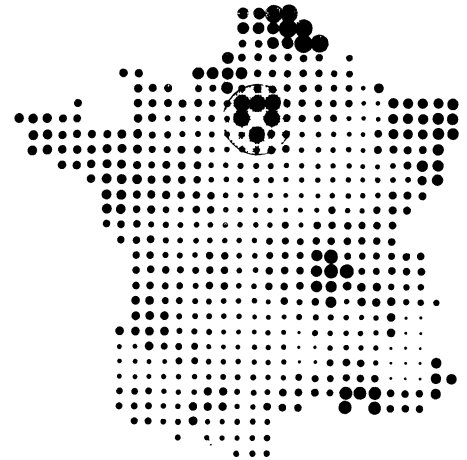


Fig. 4

LA CARTOGRAPHIE STATISTIQUE AUTOMATIQUE

Une analyse de la cartographie statistique moderne montre que la grande majorité des cartes représente des quantités par zones. Dans la plupart des cas en effet, une analyse géographique peut être ramenée à cette définition et quelle que soit la nature de l'unité décomptée le problème graphique est semblable : il s'agit de construire un histogramme à trois dimensions. Mais si la transcription de l'espace géographique est aisée, quelle forme graphique faut-il choisir pour représenter les quantités ? Nous étudierons donc : I) le choix d'une formule II) l'automatisation de cette formule.

I - LE CHOIX D'UNE FORMULE

La représentation d'une série quantitative traduit la "variation de distance" entre les catégories, variation exprimée par les quantités.

Lorsque les catégories sont alignées, c'est l'histogramme de répartition des statisticiens (fig.1). Il révèle des paliers qui permettent de "réduire" la liste des catégories à quelques termes : les groupements caractérisés par de faibles "distances quantitatives" (fig.2)

Lorsque les catégories construisent un réseau, l'histogramme devient un relief à trois dimensions (fig.3) qui doit révéler les groupes résultant non seulement des "distances quantitatives" mais aussi des "distances géographiques". Ces groupements ne correspondent pas nécessairement à ceux révélés par l'histogramme et des régions homogènes peuvent résulter de la confusion intime de catégories situées aux deux extrémités de l'histogramme (par exemple fig.9).

En conséquence :

I - seul un histogramme à trois dimensions, un "relief", peut transcrire convenablement des quantités par zones.

THE AUTOMATICAL STATISTICAL CARTOGRAPHY

An analysis of modern statistical cartography shows that the majority of maps represent quantities per zones. In most cases a geographical analysis can be reduced to this definition and whatever the nature of the geographical unit is, the graphical problem is similar : the solution is to draw a histogram in three dimensions. But if the transcription of a geographical area is easy, what graphical form should be chosen to represent quantities ? Let us then study : 1) the choice of a formula, II) the automatization of this formula.

I - THE CHOICE OF A FORMULA

The representation of a quantitative series shows the "distance variation" between categories ; these variations are given by quantities. When the categories are aligned together we have the statisticians' histogram distribution (fig.1). It reveals levels which enable the list of categories to be reduced to a few terms : groups characterized by small "quantitative distances" (fig. 2).

When categories build up a network, the histogram forms a three dimensional "relief" (fig. 3) which should reveal the groups resulting not only from "quantitative distances", but also from "geographical distances". These groups do not necessarily correspond to those shown by the histogram and some homogeneous regions may result from the intimate confusion of categories located at the two ends of the histogram (for ex. fig.9).

In consequence

I - only a histogram in three dimensions, a "relief" can properly transcribe quantities by zones.

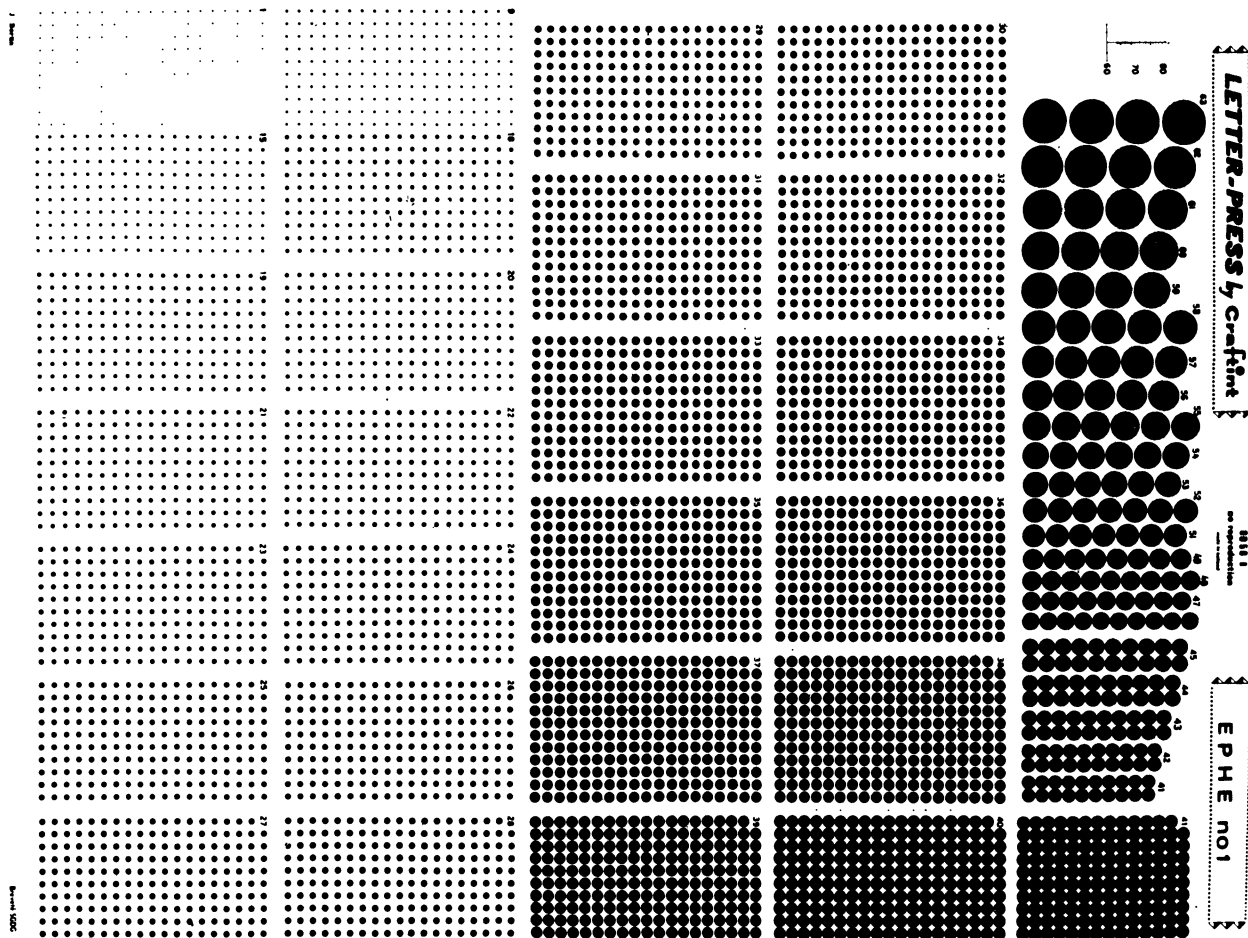


Fig.5

2 - comme l'histogramme de répartition, le relief doit d'abord représenter l'information exhaustive (c'est-à-dire non réduite à quelques paliers préconçus), après quoi, et seulement après, l'observateur peut découvrir les paliers contenus dans l'information. Les paliers sont le but de la représentation et non le moyen, ce qui exclue toute formule comportant un choix préalable de paliers.

3 - l'oeil doit pouvoir dégager de l'ensemble de la carte les zones différentes et les zones semblables, ce qui exclue toute représentation interdisant la lecture d'ensemble des paliers.

4 - les différences doivent être quantitatives et mesurables visuellement sans l'aide de nombres écrits. On démontre que parmi toutes les formules graphiques possibles, seul le semis régulier de cercles croissants (fig.4) peut à la fois être exhaustif et admettre la perception zonale, la perception quantitative et la lecture d'ensemble. La longueur de la variation sensible est considérable car cette formule combine deux perceptions convergentes : la différence de taille entre les points et la différence de valeur entre les zones. Elle permet de percevoir des quantités absolues et des densités et

2 - as the histogram of distribution, the "relief" must represent first the exhaustive information (this is to say not reduced to a few preconceived levels). After which and only after, the observer can discover the levels contained inside the information.

3 - the eye should be able to distinguish from the map as a whole the non similar zones and the similar ones, which excludes any representation preventing the reading of the levels as a whole.

4 - the differences must be quantitative and visually measurable without the aid of written numbers. It is demonstrated that among all the possible graphical formulae, only the regular arrangement of growing circles can be at the same time exhaustive and permit zonal perception, quantitative perception and a reading of the whole.

The length of noticeable variation is large, because this formula combines two converging perceptions : difference of size between circles (dots) and difference of value between zones. It allows perception of



Fig. 6

apporte une information originale quel que soit le niveau de lecture adopté par l'observateur. A ces caractères sémantiques il faut ajouter des propriétés fonctionnelles : le semis régulier autorise les plus fortes réductions (microfilmage, fig.10) et se prête remarquablement à l'automatisation.

L'établissement de l'échelle, c'est à dire de la correspondance entre la taille des cercles et la série statistique doit être très précis. Il est possible de dessiner des cercles proportionnels à chacune des quantités. Mais la séparation de deux tailles devient inutile lorsqu'elle tombe au-dessous du seuil de sensibilité. On peut donc construire la série nécessaire et suffisante : la gamme naturelle des tailles croissantes. Elle est universelle. Sa longueur est définie par l'écart au-dessous duquel toute taille intermédiaire devient inutile, et par la progression géométrique qui détermine des écarts sensibles réguliers et constants.

Cette gamme étant fixée (fig.5), le calcul de l'échelle devient très aisé. Pour bénéficier de la plus grande séparation visuelle, il suffit de faire correspondre les limites de la série quantitative avec les limites de la gamme. Les cercles intermédiaires définissent les coupures statistiques que des tables de correspondance donnent immédiatement.

II - L'AUTOMATISATION DE LA CARTOGRAPHIE STATISTIQUE

Cette formule peut être dessinée à la main. Elle peut être semi-automatique grâce aux planches préfabriquées. Elle peut enfin être automatisée. Notons que l'automatisation n'a pas été le but de nos recherches mais l'aboutissement logique des expériences qui tiennent compte

- de la généralité du problème et de son économie,
- des caractéristiques de la formule graphique efficace,
- de la masse documentaire à exploiter et de la rapidité requise.

Pour réaliser cette automatisation il faut disposer d'un ordinateur, d'un programme et d'une imprimante spéciale.

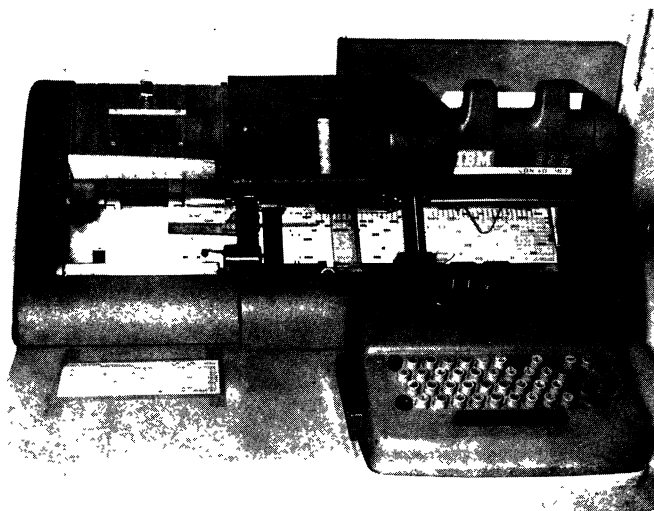


Fig.7

absolute quantities and of densities and gives original information whatever the level of reading adopted by the observers. To these semantical characteristics, we must add functional qualities : the regular arrangement allows bigger reductions (microfilming, fig.10) and can be remarkably easily used with automatization.

The choice of a scale, that is to say the correspondence between the size of the circles and the statistical series must be very precise. It is possible to draw proportional circles for each one of the quantities. But the distinction of two sizes becomes pointless when it falls below the threshold of comprehensibility. So, we can build up the necessary and sufficient series : the natural scale of growing sizes. It is universal. Its length is defined by the distance below which any intermediate size becomes useless and by the geometrical progression which determines comprehensible, regular and constant distances.

This scale being fixed, (fig.5) the calculation of the scale becomes very easy. To benefit from the biggest visual separation, it is sufficient to make the limits of the quantitative series correspond with the limits of the scale. The intermediate circles define statistical cuts which are given immediately by the correspondence tables.

II - THE AUTOMATISATION OF STATISTICAL CARTOGRAPHY

This formula can be drawn by hand. It can be semi-automatised by using ready made tables. It can even be automatised. It should be noticed that automatization has not been the aim of our research, but the final and logical issue of experiment which takes into consideration :

- the generality of the problem and its economy ;
- the characteristics of an efficient graphical formula ;
- the quantity of the documentation which is to be used and the speed required.

To achieve this automatization, a computer, a program and a special printer are required.

DEMANDES D'EMPLOI NON SATISFAITES AU 1ER SEPTEMBRE 1965
POUR 10 000 HABITANTS ACTIFS

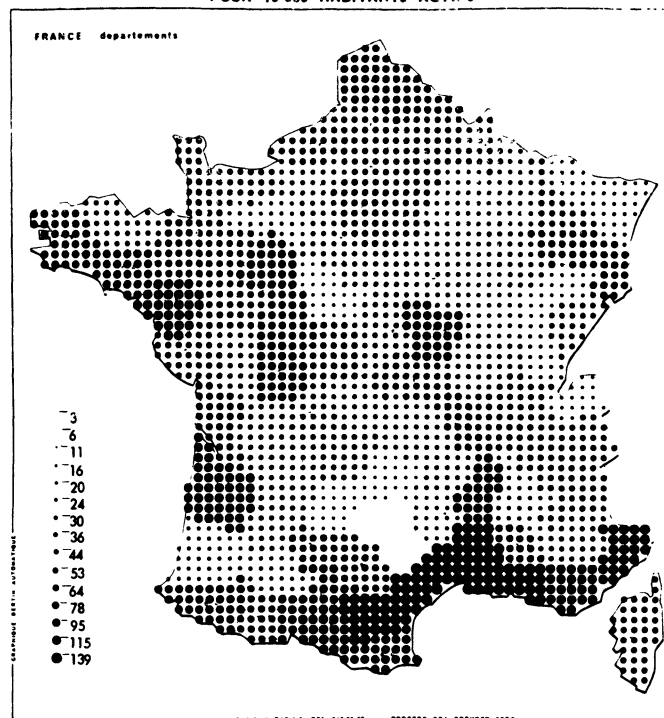


Fig. 8

L'imprimante peut être une tabulatrice, un disque lumitype, une table traçante, un écran cathodique ou une simple machine à écrire modifiée. Nous avons retenu cette dernière solution : un système IBM 870 (fig.6) modifié et équipé des signes de la gamme naturelle et d'un interlignage égal à l'écartement.

Il est commandé par un jeu de cartes perforées fourni par l'ordinateur et lu par une lectrice 836 (fig.7). Pour fournir ce jeu, l'ordinateur est instruit de la série statistique et d'un programme qui contient les constantes et commande les opérations à effectuer.

Le programme contient :

a - la correspondance entre les "adresses" de l'information (les numéros de zones) et les "adresses" en X et en Y des points de la carte. C'est une constante pour un réseau géographique donné.

b - la correspondance entre la gamme naturelle (constante) et la série statistique, après analyse automatique de la série, qui définit l'échelle et élimine éventuellement les "queues".

Le programme permet de plus :

c - de recevoir plusieurs séries et d'effectuer toutes opérations : somme, différence ou rapport de séries entre elles.

d - de définir une échelle commune à plusieurs séries.

e - de diviser les nombres par la surface des zones et calculer la densité.

OMBRE DE MEDECINS POUR 100.000 HABITANTS

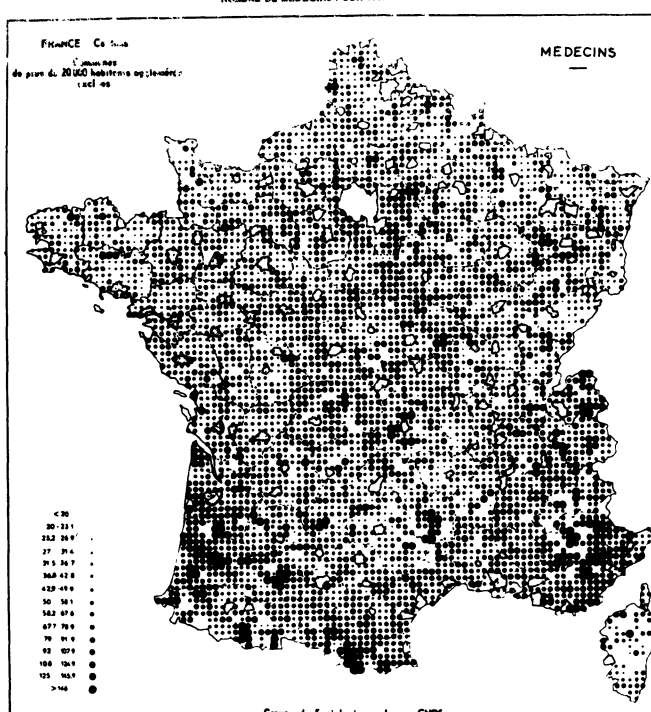


Fig. 9

The printer can be a tabulator, a lumitype disc, a plotter, a visual display unit, or a simple modified type-writer. We have chosen this last solution : an I.B.M. 870 system (fig.6) modified and equipped with the signs of a natural scale and of a space between two lines equal to the distance of dots.

It is fed by a set of punched cards produced by the computer and read by a I.B.M. 836 punched cards reader (fig.7).

To obtain this set of punched cards the statistical series is fed to the computer together with a program which contains the constants and which control the operations to be done.

The program contains :

a) the correspondence between "locations" of the information (the numbers of the zones) and the "locations" in X and in Y of the points of the map. It is constant for a given geographical network.

b) the correspondence between the natural scale (constant) and the statistical series, after automatic analysis of the series which defines the scale and eliminates if necessary "tails".

The program further enables

c) several series to be undertaken and all operations to be effected : sum, difference or relation of series between one another.

d) a common scale for several series to be defined.

e) the numbers to be divided by surface of zones and

f - de sortir "en clair" sur une feuille d'accompagnement les éléments de la légende, de l'échelle, des hors-échelles (ou "queues") et les caractéristiques mathématiques de la série.

Pour rédiger une carte

1 - on instruit l'ordinateur du programme, des séries statistiques et d'une carte perforée de contrôle.

2 - l'ordinateur fournit un jeu de cartes perforées et la feuille d'accompagnement.

3 - ce jeu est introduit dans la "lectrice" qui commande l'imprimante.

L'impression peut se faire sur papier vierge ou sur un fond de carte pré-imprimé. L'imprimante IBM crée une carte de France (départements) de 90 zones (fig.8) en 14 minutes; une carte cantonale de 3000 zones (fig.9) en 40 minutes. L'écran cathodique réalisera ces documents en quelques secondes. Toutes les opérations nécessaires peuvent être assurées par une bonne dactylographe.

Pour utiliser normalement la cartographie statistique automatique, il suffit d'instruire l'ordinateur de toutes les séries originales (mais seulement de ces séries). Toute carte résultant de ces séries, d'une somme, d'une différence ou d'un quelconque rapport de séries entre elles sort automatiquement sur simple commande. En conclusion, la mise en oeuvre de l'énorme matériel statistique progressivement disponible ne peut devenir que massive et immédiate. Sa réduction doit faire appel aux mathématiques et à la représentation graphique, hors desquelles il n'y a pas d'intégration possible. L'automatisation de la cartographie, associée à la documentation automatique est donc inéluctable. Les responsables en ont besoin pour appuyer leurs décisions sur l'ensemble de l'information moderne.

the density to be calculated.

f) to give "clearly" on a accompanying sheet the data concerning the legend, the scale, the values off scale (or "tails") and the mathematical characteristics of the series.

To draw-up a map :

1 - the program, the statistical series and a controlling punched card, is fed to the computer.

2 - the computer produces a set of punched cards and the accompanying sheet.

3 - the set of punched cards is fed into the reader which feeds the printer.

The printing can be done on blank paper or on a paper with preprinted map. The IBM printer produces a map of France (departments) of 90 zones (fig.8) in 14 minutes; a map of 3000 zones in 40 minutes. The visual display unit produces these documents in a few seconds. All the necessary operations can be done by a good typist.

To use automatic statistical cartography efficiently, it is sufficient to feed the computer with all the original series (but only these series). Any map resulting from these series, of a sum, of a difference or of any relation of series between themselves, can automatically be produced at will.

In conclusion the handling of the vast statistical material now available must be undertaken on a massive scale and immediately. Its reduction requires the use of mathematics and graphical representation without which there can be no possible integration. The automation of cartography, associated with automatic documentation, is therefore inescapable. Those responsible need it to base their decisions on the fullest modern information.

Jacques BERTIN



Fig. 10